Chasing the Rubicon?

Michael C. Frank

Department of Psychology, Stanford University

Thanks to Dominic Massaro, Ellen Markman, and Mark Seidenberg for helpful comments.

Please address correspondence to Michael C. Frank, Department of Psychology, Stanford University, 450 Serra Mall (Jordan Hall), Stanford, CA, 94305, tel: (650) 724-4003, email: mcfrank@stanford.edu.

> Where, then, is the difference between brute and man? What is it that man can do, and of which we find no signs, no rudiments, in the whole brute world? I answer without hesitation: The one great barrier between the brute and man is language. Man speaks, and no brute has ever uttered a word. Language is our Rubicon, and no brute will dare to cross it.
>
> Max Müller, Lectures on the Science of Language (1861), p. 360

Introduction

What makes humans different from other animals? Müller called language our "rubicon"—the unbreakable barrier between man and beast. Modern psychologists and linguists know a lot more about language than Müller did; nevertheless, many current hypotheses are not that different than Müller's. Contemporary theorists tend to give the Rubicon a name, whether it is recursion (Hauser, Chomsky, & Fitch, 2002), symbolic thought (Deacon, 1998), reference (Bloom, 2004), or some other. These names identify the construct within language that is claimed to be unique to humans. In Hauser et al. (2002)'s influential nomenclature, this construct constitutes the "faculty of language, narrow." In other words, the rubicon.

From this perspective, the growing number of reports about language-trained animals is important and potentially problematic. If there really is a dividing line between human beings and other animals, then what are we to make of the growing menagerie that includes Kanzi the Bonobo (Savage-Rumbaugh, 1994), Alex the Parrot (Pepperberg & Pepperberg, 2009), various dolphins (Herman, Richards, & Wolz, 1984), or the pack of dogs including Rico (Kaminski, Call, & Fischer, 2004), Bailey (Griebel & Oller, 2012), and now Chaser (Pilley & Reid, 2011)? On such a view, these cases must either be miraculous discoveries or trivial mistakes.

In the case that an analyst can declare the behavior to be "non-linguistic," the language-trained animal is showing a trivial party trick—perhaps one that reflects the malleability of the animal and the skill of the trainer—but not a scientific discovery that constrains theories of language or human nature. On the other hand, if the analyst cannot reject the importance of the behavior, then it is a miraculous discovery and a hypothesis about the nature of language must be rejected. The second outcome is not common.

An example of the failure of this binary strategy comes from the recursion debate. Hauser et al. (2002) proposed "recursion" as their rubicon. As a response, Gentner, Fenn, Margoliash, and Nusbaum (2006) claimed to show that starlings, an unrelated species, could discriminate center-embedded strings from tail-recursive strings, a key test case for one particular formal definition of recursion. But—presuming for a moment that we accept this evidence (cf. Corballis, 2007)—what is the inference we must make? On the rubicon viewpoint, we must reject that recursion of this sort is constitutive of the human-unique portions of the language faculty. Such a rejection would be a major discovery about the nature of human language.

But the truth is likely to be both more banal and more complicated. Starlings are not particularly closely related to humans, and they certainly don't share any common ancestors that have the same pattern learning abilities. Whatever they are doing to discriminate strings may have little to do computationally with human linguistic abilities. Evn as a case of convergent evolution, the claim is quite weak: Starlings don't show the same behavior as humans. They simply share—putatively—the cognitive ability to process strings in the same formal complexity class (cf. Perruchet & Rey, 2005). So it's hard to see why—absent an interpretation of the behavior as rubicon-crossing—an interesting avian ability should reveal much about human language.

So why don't we reject the rubicon? As an alternative, we could begin by acknowledging that human language is different in kind not because of a single computational feature, but because it is the sum of many smaller differences from attested animal communication systems (Pinker & Jackendoff, 2005). This multifactorial view is unsatisfying and under-constrained, but more likely to be true. With this context, let's turn to Chaser, a fascinating study that could be more scientifically illuminating were it not quite so preoccupied with rubicon-crossing.

Chaser: Unlocking the Genius of the Dog Who Knows a Thousand Words

Chaser is a Border Collie raised by John Pilley, a retired psychology professor at Wofford College. As Pilley's (2013b) engaging narrative describes, by the time of writing, Chaser had mastered the names of over a thousand objects. Upon hearing one of these names, Chaser could fetch the matching object from a large comparison set with high accuracy and under a variety of quite stringent testing conditions: in public, with other speakers, and in controlled setups with no opportunity for social cueing. Ample video documentation attests to this achievement. Given that Chaser's skill is consistent with—though quantitatively great than—that of other carefully-trained dogs (Kaminski et al., 2004; Griebel & Oller, 2012), I personally don't see any reason to doubt the evidence.

It's hard to imagine anyone more well-suited to training Chaser than Pilley, who retired after a career spent teaching and studying animal learning. From very early in Chaser's life, Pilley devoted 4–5 hours a day to training word meanings through a series of carefully-designed procedures. It's abundantly clear that Pilley cares deeply about Chaser. This devotion—along with Chaser's corresponding attachment—form the motivational core that would allow for such a remarkable achievement.

Of course, one potential critique of this work has to do with the quantity of training that Chaser received. In an earlier commentary on canine word learning, Bloom (2004) notes that it took the dog Rico nine years to master 200 words, while human children learn thousands or tens of thousands during that time. I am not quite sure how useful such quantitative comparisons are, going forward; this de-emphasis on quantity alone is perhaps an important contribution of Pilley's work with Chaser. How different is 1000 words—a milestone that many children reach during their third year (Mayor & Plunkett, 2011)—from the 40,000 I am told by various online tests that I know?

Aside from this question of numerical line-drawing (more rubicons to cross!), there is the question of whether these numbers reflect differences in mechanism, or merely in experience and learning rate. In human language acquisition, there is an increasing appreciation of the immense amount of language input children receive—one back-of-the-envelope calculation suggests up to a million words a month in some case, leading common words to be heard hundreds, thousands, or tens of thousands of times before they are first produced (Frank, Tenenbaum, & Gibson, 2013). And the amount of language that children hear appears strongly related to their vocabulary size and facility with language (Hart & Risley, 1995; Hoff, 2003; Weisleder & Fernald, 2013). For children, and almost certainly for dogs as well, hearing more language—ideally in supportive, clear situations—leads to learning more words. The dogs probably learn slower, but that alone isn't evidence for much of anything about the type of learning they are doing.

In addition to learning to recognize a large number of words, Chaser showed several other behaviors of interest. First, he was able to learn by exclusion, demonstrating success in a "mutual exclusivity" experiment of the type commonly used in studying word learning by human children (e.g. Markman & Wachtel, 1988; Bion, Borovsky, & Fernald, 2013). Second, he was able to master taxonomic categories for objects, responding correctly to "toy," "ball," and "frisbee" as category labels that covered several objects that also had their own names. Finally, and perhaps most impressively, Chaser learned to respond appropriately to multi-word commands, distinguishing "paw lamb" (touch with your paw) from "nose lamb" (push with your nose) and even learning to respond correctly to commands like "To Jeffrey... take decoy" (Pilley, 2013a). I'll discuss learning by exclusion and multi-word combination below, as well as the nature of Chaser's words.

Learning by Exclusion

The simplest "mutual exclusivity" experiment goes like this: a child is presented with two objects, one familiar and one novel, and is asked to give the experimenter the "dax" (a novel label that the child presumably has never heard before). Across dozens of studies, children as young as 14 – 18 months show evidence of this kind of behavior (Halberda, 2003; Markman, Wasow, & Hansen, 2003), although they do not tend to retain the link between the novel word and novel object until significantly later (Horst & Samuelson, 2008; Bion et al., 2013).

While the phenomenon of mutual exclusivity is clear and highly replicable, the psychological sources of children's behavior remain obscure. Some accounts have posited that children reason from basic principles either about the structure of the lexicon (Markman, 1990) or about novel stimulus mappings (Mervis & Bertrand, 1994); others have proposed explanations of the same behavior in terms of pragmatic reasoning about language use in context (Clark, 1990). Still others, myself included, have argued that this behavior could in principle arise from general principles of statistical inference (Regier, 2005; Frank, Goodman, & Tenenbaum, 2009). In principle, these explanations themselves are not mutually exclusive, and could support the same behavior to different degrees at different times in development (Lewis & Frank, 2013).

Given this situation, I'm not sure what we can infer from the fact of Chaser's learning by exclusion. The evidence—as I see it, at least—doesn't warrant any reverse inference from an animal displaying learning by exclusion to a particular explanation of *why* they are showing that behavior. Chaser's very slow reaction times in mutual exclusivity tasks—a fascinating observation that is tossed in as a device to add narrative tension about his success—could perhaps provide a clue, but we don't get any quantitative details.

In addition, Chaser's display of learning by exclusion is somewhat problematic. Rico the dog was also able to succeed in this type of experiment, but, as Markman and Abelev (2004) note, Kaminski et al. (2004) had no control condition testing whether Rico succeeded because of a baseline preference for novelty. Perhaps Rico would simply have retrieved a novel object even if there were no label, just because it was more attractive. Pilley repeats the same experiment that Kaminski et al. (2004) did, but adds a control condition in which Chaser is asked to fetch familiar competitors rather than the novel object. Unfortunately, this control condition doesn't address the underlying issue raised by Markman and Abelev. Chaser could still have a novelty preference that led him to choose the novel referent in the absence of a familiar label, if this preference were weaker than the learned labels of the familiar objects.

Chaser's "Grammar"

Beyond the sheer quantity he aimed for, Pilley's experiments on multi-word utterances are perhaps the most novel aspect of his work with Chaser. As reported in Pilley (2013a), Chaser was able to comprehend sentences of the form "to frisbee take ball" (meaning, take the ball to the frisbee). Chaser performed above chance (75% correct) in executing instructions of this type, using the names of objects that had not been specifically trained in this task.

A few aspects of this set of experiments stand out. First, the odd word order Pilley chooses is not an accident: It is optimized to Chaser's cognition. In pilot work, Pilley attempted to use English word order but found that Chaser typically took the object named by the final word in his mouth, even though it was meant to be the location (e.g., "take the ball to the frisbee" implies that you should pick up the ball, not the frisbee). This learnability finding suggests that Chaser's cognitive limitations—in this case, perhaps a recency bias in memory—shape his ability to learn particular regularities. Suggestions about the effects of cognitive limitations on language structure have been made regarding human language as well (Christiansen & Chater, 2008), so it would be of substantial interest to understand what kinds of regularities Chaser can and cannot learn (and perhaps how those relate to canine sequence processing abilities). Such studies might be more informative than the mere fact of success.

Second, although Chaser's word-order learning is a substantial accomplishment, it's worth noting that Pilley's behaviorist vocabulary isn't really equipped to describe the findings that he gets. He writes that "as the farmer alters and labels the goals of the herding, such as herding the sheep toward or away from the farmer or to different locations, the dog infers that the name of the location specifies a given location to which the dog must herd the sheep without the farmer; thereafter, the dog needs only a single verbal command to take the sheep to the designated location." (Pilley, 2013a). This vocabulary doesn't have a way to describe how Chaser represents abstractions like "location," or how he maps them to positions in a string of words. In that sense, it is strongly reminiscent of the circular accounts given in previous behaviorist analyses of language (Skinner, 1957).

Are Chaser's Words Just Associations?

The question of whether *human children's* words are "mere associations" has been the subject of a substantial debate both historically and in recent years (e.g. Sloutsky, Lo, & Fisher, 2001; Waxman & Gelman, 2009). The same thing can be asked about Chaser's words. This debate often tends to be more circular than informative. We might instead ask two related questions: first, what is the nature of Chaser's representation of words, and second, what are the computations (inferences) that he can use them in.

My own guess is that words are represented by humans as associations between linguistic forms and concepts. These associative links are long-term memories that obey the regularities governing such associations (Bahrick, Bahrick, Bahrick, & Bahrick, 1993; Frank et al., 2013). They may gain some privilege due to modality: associations with words are highly salient, and words themselves are represented in much more detail than, say, arbitrary tones. But I'd argue that what makes our language special is not the words themselves, but the inferences we make about the *use* of words.

Human speakers use words to refer to the world and to communicate about both internal and external content. What's more, even in early childhood, children seem to appreciate the nature of this usage: They are able to make a wide variety of powerful inferences from the fact and circumstances of language use (Baldwin, 1995; Vouloumanos, Onishi, & Pogue, 2012). This distinction between the long-term storage of words and the inferences that can be made from their use in context is critical to recent computational accounts of word learning (Frank et al., 2009; McMurray, Horst, & Samuelson, 2012). Distinguishing learning and infererence also helps to explain a number of empirical facts about children's behavior in word learning experiments, like the rich-get-richer nature of word learning or the previously-mentioned ability to make mutual exclusivity inferences without later retaining the target word.

Seen from this perspective, Chaser might have similar representations of word meaning to human children, at least for the concrete nouns and noun category labels that he knows. What still bears exploration is the extent to which he can make inferences about the social role that labels play. I would love to see his performance in classic tasks by Baldwin (1991), in which cues to the speaker's intention are pitted against salience and temporal contiguity. Or, further afield, he could even be tested in looking-time paradigms like those of Vouloumanos et al. (2012) to test expectations about the communicative functions of language. Such explorations would shed light on the degree to which Chaser is able to use words flexibly as a marker of social communication.

Conclusions

Chaser is a touching book that reports on the product of an astonishing amount of sustained effort by Pilley. Certainly it feels like an endpoint that may not soon be repeated in canine language training. On the other hand, I am not sure what it tells us about human language. I say that reluctantly, since there is a lot to like about this work. But its primary contribution to the discussion of human language is to provide (further) evidence against a series of rubicons—learning by exclusion and multi-word combination, to name two.

If we reject this kind of dichotomous thinking and instead consider the simultaneous contributions of many different abilities to human language, the interest in Chaser will be not *that* he can show particular behaviors, but *how* he performs them. I would have loved to see experiments that show the circumstances under which Chaser *can't* learn a word, to test hypotheses about the necessary ingredients for learning. For example, what would his abilities be in incidental learning (Akhtar, 2005)? How would he perform in a range of intention reading and social-cue combination tasks (Baldwin, 1991; Hollich, Hirsh-Pasek, & Golinkoff, 2000)? Unfortunately the book and accompanying scientific papers are silent on Chaser's limitations.

Chaser is admirable attempt to jump across the rubicon. But if we abandon the notion of rivers to be crossed, and move instead towards a multifactorial view of what makes human language unique, we need to acknowledge that what once seemed a river may be a wide, muddy swamp instead. Exploring this landscape may require more slogging and less jumping.

References

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science*, *8*, 199–209.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, *4*, 316–321.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*, 874–890.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 131–158). Hillsdale, NJ: Erlbaum.

Bion, R. A., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, *126*, 39–53.

Bloom, P. (2004). Can a dog learn a word? *Science*, *304*, 1605–1606.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*, 489–509.

Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, *17*, 417–431.

Corballis, M. C. (2007). Recursion, language, and starlings. *Cognitive Science*, *31*, 697–704.

Deacon, T. W. (1998). *The symbolic species: The co-evolution of language and the brain.* WW Norton & Company.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, *20*, 578–585.

Frank, M. C., Tenenbaum, J. B., & Gibson, E. (2013). Learning and long-term retention of large-scale artificial languages. *PloS ONE*, *8*, e52500.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, *440*, 1204–1207.

Griebel, U., & Oller, D. K. (2012). Vocabulary learning in a yorkshire terrier: slow mapping of spoken words. *PLoS ONE*, *7*, e30182.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, *87*, B23–B34.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young american children.* Paul H Brookes Publishing.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*, 1569–1579.

Herman, L. M., Richards, D. G., & Wolz, J. P. (1984). Comprehension of sentences by bottlenosed dolphins. *Cognition*, *16*, 129–219.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*, 1368–1378.

Hollich, G. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, 1–135.

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*, 128–157.

Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog: evidence for "fast mapping". *Science*, *304*, 1682–1683.

Lewis, M., & Frank, M. C. (2013). Modeling disambiguation in word learning via multiple probabilistic constraints. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society.*

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*, 57–77.

Markman, E. M., & Abelev, M. (2004). Word learning in dogs? *Trends in Cognitive Sciences*, *8*, 479–481.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121–157.

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*, 241–275.

Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from cdi analysis. *Developmental Science*, *14*, 769–785.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*, 831.

Mervis, C. B., & Bertrand, J. (1994). Acquisition of the novel name–nameless category (N3C) principle. *Child Development*, *65*, 1646–1662.

Pepperberg, I. M., & Pepperberg, I. M. (2009). *The Alex studies: Cognitive and communicative abilities of grey parrots.* Harvard University Press.

Perruchet, P., & Rey, A. (2005). Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin & Review*, *12*, 307–313.

Pilley, J. W. (2013a). Border collie comprehends sentences containing a prepositional object, verb, and direct object. *Learning and Motivation*, *44*, 229–240.

Pilley, J. W. (2013b). *Chaser: Unlocking the genius of the dog who knows a thousand words.* Houghton Mifflin Harcourt.

Pilley, J. W., & Reid, A. K. (2011). Border collie comprehends object names as verbal referents. *Behavioural Processes*, *86*, 184–195.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition*, *95*, 201–236.

Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, *29*, 819–865.

Savage-Rumbaugh, E. S. (1994). *Kanzi: The ape at the brink of the human mind.* Hoboken, NJ: Wiley.

Skinner, B. F. (1957). *Verbal behavior.* Chicago, IL: Copley Publishing Group.

Sloutsky, V. M., Lo, Y.-F., & Fisher, A. V. (2001). How much does a shared name make things similar? linguistic labels, similarity, and the development of inductive inference. *Child Development*, *72*, 1695–1709.

Vouloumanos, A., Onishi, K. H., & Pogue, A. (2012). Twelve-month-old infants recognize that speech can communicate unobservable intentions. *Proceedings of the National Academy of Sciences*, *109*, 12933–12937.

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13*, 258–263.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*, 2143–2152.